

LIBRARY
OF THE
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

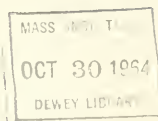
Bayesian Analysis of the Independent Multi-

Normal Process--Neither Mean Nor

Precision Known

Albert Ando and G.M. Kaufman

Revised, September, 1964



41-63

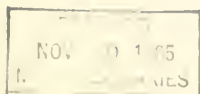
The contribution by Ando to this paper is partially supported by a grant from the National Science Foundation. The authors wish to thank the referees for helpful suggestions and Jim Martin for proofreading early drafts. We gratefully acknowledge valuable comments and criticisms given to them by members of the Decisions Under Uncertainty Seminar, conducted by Professor Howard Raiffa at the Harvard Business School.

1022

M414

14 4 83

Dewey



Bayesian Analysis of the Independent Multinormal Process--Neither

Mean nor Precision Known

By Albert Ando and G.M. Kaufman

SUMMARY

Under the assumption that neither the mean vector nor the variance-covariance matrix are known with certainty, the natural conjugate family of prior densities for the multivariate Normal process is identified. Prior-posterior and preposterior analysis is done assuming that the prior is in the natural conjugate family. A procedure is presented for obtaining non-degenerate joint posterior and preposterior distributions of all parameters even when the number of objective sample observations is less than the number of parameters of the process.

1. Introduction

In this paper we develop the distribution theory necessary to carry out Bayesian analysis of the multivariate Normal process as defined in section 1.1 below when neither the mean vector nor the variance-covariance matrix of the process is known with certainty. The development here generalizes Raiffa's and Schlaifer's treatment of the multivariate Normal process as done in Part B, Chapter 12 of [5], in which it is assumed that the variance-covariance matrix is known up to a particular multiplicative constant. We drop this assumption here.

In section 1 we define the process, identify a class of natural conjugate distributions, and do prior -- posterior analysis. The conditional and unconditional sampling distributions of some (sufficient) statistics are presented in section 2. In particular, we prove that the distribution of the sample mean vector marginal with respect to the sample variance-covariance matrix, to the process mean vector, and to the process variance-covariance matrix is multivariate Student whenever the prior is in the natural conjugate family of distributions. We then use the results of sections 1 and 2 to do preposterior analysis in section 3.

- 2 -

We also show in section 3 that Bayesian joint inference--finding joint posterior and preposterior densities of the mean vector and the variance-covariance matrix--is possible even when classical joint inference is not, i.e. when the number of objective sample observations is less than the number of distinct elements of the mean vector and variance-covariance matrix of the process

Geisser and Cornfield [3] and Tiao and Zellner [7] analyze the multivariate Normal process and multivariate Normal Regression process respectively under identical assumptions about the state of knowledge of the parameters of the process. Their presentations differ from that given here in three respects: first, following the lead of Jeffreys [4], both sets of authors assume that the joint prior on the mean vector and variance-covariance matrix is a special (degenerate) case of a natural conjugate density; second, here we find sampling distributions unconditional as regards the parameters of the process and do preposterior analysis; third, by doing the analysis for the complete natural conjugate family, we are able to provide a procedure for deriving joint posterior and some joint preposterior distributions of all parameters under conditions mentioned in the paragraph immediately above.

1.1 Definition of the Process

As in [1], we define an r -dimensional Independent Multinormal process as one that generates independent $r \times 1$ random vectors $\underline{\tilde{x}}^{(1)}, \dots, \underline{\tilde{x}}^{(j)}, \dots$ with identical densities

$$f_N^{(r)}(\underline{x} | \underline{\mu}, \underline{h}) = (2\pi)^{-\frac{1}{2}r} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^t \underline{h}(\underline{x}-\underline{\mu})} |\underline{h}|^{\frac{1}{2}} \quad (1)$$

$$\begin{aligned} -\infty < \underline{x} < \infty, \\ -\infty < \underline{\mu} < \infty, \\ \underline{h} \text{ is PDS} \end{aligned}$$

1.2 Likelihood of a Sample

The likelihood that the process will generate n successive values $\underline{x}^{(1)}, \dots, \underline{x}^{(j)}, \dots, \underline{x}^{(n)}$ is

$$(2\pi)^{-\frac{1}{2}rn} e^{-\frac{1}{2}\Sigma(\underline{x}^{(j)} - \underline{\mu})^t \underline{h}(\underline{x}^{(j)} - \underline{\mu})} |\underline{h}|^{\frac{1}{2}n} \quad (2)$$

If the stopping process is non-informative, as defined in [1], this is the likelihood of a sample consisting of n observations $\underline{x}^{(1)}, \dots, \underline{x}^{(j)}, \dots, \underline{x}^{(n)}$.

When neither $\underline{h}$ nor $\underline{\mu}$ is known, we may compute these statistics:

$$\underline{m} \equiv \frac{1}{n} \Sigma \underline{x}^{(j)} \quad , \quad v \equiv n-r \text{ (redundant)} \quad , \quad (3a)$$

and

$$\underline{v} \equiv \Sigma(\underline{x}^{(j)} - \underline{m})(\underline{x}^{(j)} - \underline{m})^t \quad . \quad (3b)$$

It is well known[†] that the kernel of the joint likelihood of $(\underline{m}, \underline{v})$ is, provided $v > 0$,

$$e^{-\frac{1}{2}n(\underline{m}-\underline{\mu})^t \underline{h}(\underline{m}-\underline{\mu})} |\underline{h}|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\text{tr } \underline{h} \underline{v}} |\underline{h}|^{\frac{1}{2}(v+r-1)} \quad , \quad (4a)$$

the kernel of the marginal likelihood of $\underline{m}$ is, provided $v > 0$,

$$e^{-\frac{1}{2}n(\underline{m}-\underline{\mu})^t \underline{h}(\underline{m}-\underline{\mu})} |\underline{h}|^{\frac{1}{2}} \quad , \quad (4b)$$

and the kernel of the marginal likelihood of $\underline{v}$ is, provided $v > 0$, and $\underline{v}$ is PDS,

[†]See for example, Anderson [1], Theorem 3.3.2 and pp. 154-160.

$$e^{-\frac{1}{2}\text{tr } \underline{h} \underline{V}} |\underline{h}|^{\frac{1}{2}(v+r-1)} \quad (4c)$$

Formula (4c) is the kernel of a Wishart distribution.

A random matrix $\underline{\tilde{V}}$ of dimension $(r \times r)$ will be called "Wishart distributed with parameter $(\underline{h}, v)$ " if

$$\underline{\tilde{V}} \sim f_W^{(r)}(\underline{V} | \underline{h}, v) = \begin{cases} w(r, v) |\underline{V}|^{\frac{1}{2}v-1} e^{-\frac{1}{2}\text{tr } \underline{h} \underline{V}} |\underline{h}|^{\frac{1}{2}(v+r-1)} & \text{if } \underline{V} \text{ is PDS and } v > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (4d)$$

where

$$w(r, v) \equiv [2^{\frac{1}{2}(v+r-1)r} \pi^{r(r-1)/4} \prod_{i=1}^r \Gamma(\frac{1}{2}(v+r-i))]^{-1}. \quad (4e)$$

That $(\underline{m}, \underline{V}, v)$ defines a set of sufficient statistics for $(\underline{\mu}, \underline{h})$ is shown in section 3.3.3 of [1].

We will wish to express (4a) in such a fashion that it automatically reduces to (4b) when only $(\underline{m}, n)$ is available and to (4c) when only $(\underline{V}, n)$ is available. In addition, we will wish to treat the cases that arise when $\underline{V}$ is singular. Hence we define

$$\Phi = \begin{cases} n-1 & v \leq 0 \\ 0 & v > 0 \end{cases} \quad (5a)$$

$$\underline{V}^* = \begin{cases} \underline{V} & \text{if } \underline{V} \text{ is non-singular} \\ 0 & \text{otherwise} \end{cases} \quad (5b)$$

and

$$\delta = \begin{cases} 1 & n > 0 \\ 0 & n = 0 \end{cases} \quad (5c)$$

In terms of (5a), (5b) and (5c) we may rewrite (4a), (4b) and (4c) as

$$e^{-\frac{1}{2}n(\underline{m}-\underline{\mu})^t \underline{h}(\underline{m}-\underline{\mu})} |\underline{h}|^{\frac{1}{2}\delta} e^{-\frac{1}{2}\text{tr } \underline{h} \underline{V}^*} |\underline{h}|^{\frac{1}{2}(\nu-\phi+r-1)}, \quad (4a')$$

$$e^{-\frac{1}{2}n(\underline{m}-\underline{\mu})^t \underline{h}(\underline{m}-\underline{\mu})} |\underline{h}|^{\frac{1}{2}\delta}, \quad (4b')$$

$$e^{-\frac{1}{2}\text{tr } \underline{h} \underline{V}^*} |\underline{h}|^{\frac{1}{2}(\nu-\phi+r-1)}. \quad (4c')$$

Notice that (4a) is now defined even when $\nu < 0$.

By adopting the convention that

(1) $\underline{V}^* = \underline{0}$ and $\phi=n-1$ when $\underline{V}$ is unknown or irrelevant, and

(2) $n=0$ when $\underline{m}$ is unknown or irrelevant,

the kernel (4a') reduces to (4b') in the first case and to (4c') in the second.

1.3 Conjugate Distributions of $(\underline{\mu}, \underline{h}), \underline{\mu}$ and $\underline{h}$

When both $\underline{\mu}$ and $\underline{h}$ are random variables, the natural conjugate of (4a') is the Normal-Wishart distribution $f_{NW}^{(r)}(\underline{\mu}, \underline{h} | \underline{m}, \underline{V}, n, \nu)$ defined as equal to

$$k(r, \nu) e^{-\frac{1}{2}n(\underline{\mu}-\underline{m})^t \underline{h}(\underline{\mu}-\underline{m})} |\underline{h}|^{\frac{1}{2}\delta} e^{-\frac{1}{2}\text{tr } \underline{h} \underline{V}^*} |\underline{V}^*|^{\frac{1}{2}(\nu+r-1)} |\underline{h}|^{\frac{1}{2}\nu-1} \\ = \begin{cases} f_N^{(r)}(\underline{\mu} | \underline{m}, \underline{h}n) f_W^{(r)}(\underline{h} | \underline{V}, \nu) & \text{if } n \geq 0 \text{ and } \nu > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (6a)$$

where $\underline{V}^*$ and δ are defined as in (5), and

$$k(r, \nu) = (2\pi)^{-\frac{1}{2}r} n^{\frac{1}{2}r\delta} w(r, \nu). \quad (6b)$$

If (6a) is to be a proper density function $\underline{V}$ must be PDS, $\nu > 0$ and $n > 0$ so that in this case $\delta=1$ and $\underline{V}^*=\underline{V}$ is PDS. We write the first expression in (6a) with δ and $\underline{V}^*$ so that formulas for posterior densities will generalize automatically to the case where prior information is such that one or more of these conditions hold: $\underline{V}$ is singular, $n=0$, $\nu=0$.

We obtain the marginal prior of $\underline{\mu}$ by integrating (6a) with respect to $\underline{h}$; if $\nu > 0$, $n > 0$, $\underline{V}$ is PDS, and we define $\underline{H}_V = \nu n \underline{V}^{-1}$, then

$$D(\underline{\mu}|\underline{m}, \underline{v}, n, \nu) = f_S^{(r)}(\underline{\mu}|\underline{m}, \underline{H}_\nu, \nu) \propto [\nu + (\underline{\mu}-\underline{m})^t \underline{H}_\nu (\underline{\mu}-\underline{m})]^{-\frac{1}{2}(\nu+r)} \quad (6c)$$

This distribution of $\underline{\mu}$ is the non-degenerate multivariate Student distribution defined in formula (8-26) of [5].[†]

Proof: We integrate over the region $R_{\underline{h}} \equiv \{\underline{h} | \underline{h} \text{ is PDS}\}$

$$\begin{aligned} D(\underline{\mu}|\underline{m}, \underline{v}, n, \nu) &= \int_{R_{\underline{h}}} f_N^{(r)}(\underline{\mu}|\underline{m}, \underline{h}) f_W^{(r)}(\underline{h}|\underline{v}, \nu) d\underline{h} \\ &\propto \int_{R_{\underline{h}}} e^{-\frac{1}{2}\text{tr } \underline{h}(n(\underline{\mu}-\underline{m})(\underline{\mu}-\underline{m})^t + \underline{v})} |\underline{h}|^{\frac{1}{2}(\nu+\delta)-1} d\underline{h} . \end{aligned}$$

But as the integrand in the integral immediately above is the kernel of a Wishart density with parameter $(n(\underline{\mu}-\underline{m})(\underline{\mu}-\underline{m})^t + \underline{v}, \nu+\delta)$,

$$D(\underline{\mu}|\underline{m}, \underline{v}, n, \nu) \propto [1 + (\underline{\mu}-\underline{m})^t (n\underline{v}^{-1})(\underline{\mu}-\underline{m})]^{-\frac{1}{2}(\nu+\delta+r-1)} .$$

Provided that $\nu > 0$, $\underline{v}$ is PDS and $n > 0$, $\underline{H}_\nu \equiv \nu n \underline{v}^{-1}$ is PDS, $\delta=1$, and we have (6c).

If $\nu > 0$ but $\underline{v}$ is singular, $\underline{v}^{-1}$ does not exist so neither does the marginal distribution of $\underline{\mu}$. And if $n=0$ the marginal distribution of $\underline{\mu}$ does not exist.

Similarly, we obtain the marginal prior on $\underline{h}$ by integrating (6a) with respect to $\underline{\mu}$. If $\nu > 0$, $n \geq 0$ and $\underline{v}$ is PDS, then

$$D(\underline{h}|\underline{m}, \underline{v}, n, \nu) = f_W^{(r)}(\underline{h}|\underline{v}, \nu) \propto e^{-\frac{1}{2}\text{tr } \underline{h} \underline{v}} |\underline{h}|^{\frac{1}{2}\nu-1} . \quad (6d)$$

If a Normal-Wishart distribution with parameter $(\underline{m}', \underline{v}', n', \nu')$ is assigned to $(\underline{\mu}, \underline{h})$ and if a sample then yields a statistic $(\underline{m}, \underline{v}, n, \nu)$ the posterior distribution of $(\underline{\mu}, \underline{h})$ will be Normal-Wishart with parameter $(\underline{m}'', \underline{v}'', n'', \nu'')$

[†]Cornfield and Geisser [] prove a similar result: if the prior on $(\underline{\mu}, \underline{h})$ has a kernel $|\underline{h}|^{\frac{1}{2}\nu'-1}$, $\nu' > 0$, and we observe a sample which yields a statistic $(\underline{m}, \underline{v}, n, \nu)$, $\nu > 0$, then the marginal posterior distribution of $\underline{\mu}$ is multivariate Student.

where

$$\underline{\underline{n}}' = \underline{n}' + \underline{n} \quad , \quad \delta'' = \begin{cases} 1 & n'' > 0 \\ \text{if} & \\ 0 & n'' = 0 \end{cases} \quad , \quad \underline{\underline{m}}'' = n''^{-1} (n' \underline{\underline{m}}' + n \underline{\underline{m}}) \quad , \quad (7a)$$

$$v'' = v' + v + r + \delta + \delta' - \delta'' - \phi - 1 \quad , \quad (7b)$$

$$\underline{\underline{v}}^{*''} = \begin{cases} \underline{\underline{v}}' + \underline{\underline{v}} + n' \underline{\underline{m}}' \underline{\underline{m}}'^t + n \underline{\underline{m}} \underline{\underline{m}}^t - n'' \underline{\underline{m}}'' \underline{\underline{m}}''^t & \text{if } \underline{\underline{v}}'' \text{ is PDS} \\ 0 & \text{otherwise} \end{cases} \quad . \quad (7c)$$

Proof: When $\underline{\underline{v}}'$ and $\underline{\underline{v}}$ are both PDS, the prior density and the sample likelihood combine to give the posterior density in the usual manner. When either $\underline{\underline{v}}'$ or $\underline{\underline{v}}$ or both are singular, the prior density (6a) or the sample likelihood or both may not exist. Even in such cases, we wish to allow for the possibility that the posterior density may be well defined. For this purpose, we define the posterior density in terms of $\underline{\underline{v}}'$ and $\underline{\underline{v}}$ rather than $\underline{\underline{v}}^{*}$ and $\underline{\underline{v}}^{*}$. Thus, multiplying the kernel of the prior density by the kernel of the likelihood, we obtain

$$\begin{aligned} & e^{-\frac{1}{2} n' (\underline{\underline{\mu}} - \underline{\underline{m}}')^t \underline{\underline{h}} (\underline{\underline{\mu}} - \underline{\underline{m}}')} |\underline{\underline{h}}|^{\frac{1}{2} \delta'} e^{-\frac{1}{2} \text{tr } \underline{\underline{h}} \underline{\underline{v}}'} |\underline{\underline{h}}|^{\frac{1}{2} v' - 1} \\ & \cdot e^{-\frac{1}{2} n (\underline{\underline{m}} - \underline{\underline{\mu}})^t \underline{\underline{h}} (\underline{\underline{m}} - \underline{\underline{\mu}})} |\underline{\underline{h}}|^{\frac{1}{2} \delta} e^{-\frac{1}{2} \text{tr } \underline{\underline{h}} \underline{\underline{v}}} |\underline{\underline{h}}|^{\frac{1}{2} (v + r - \phi - 1)} \\ & = e^{-\frac{1}{2} S} |\underline{\underline{h}}|^{\frac{1}{2} \delta''} e^{-\frac{1}{2} \text{tr } \underline{\underline{h}} (\underline{\underline{v}}' + \underline{\underline{v}})} |\underline{\underline{h}}|^{\frac{1}{2} (v' + v + r + \delta' + \delta - \delta'' - \phi - 1) - 1} \end{aligned} \quad (7d)$$

where

$$S \equiv n' (\underline{\underline{\mu}} - \underline{\underline{m}}')^t \underline{\underline{h}} (\underline{\underline{\mu}} - \underline{\underline{m}}') + n (\underline{\underline{m}} - \underline{\underline{\mu}})^t \underline{\underline{h}} (\underline{\underline{m}} - \underline{\underline{\mu}}) \quad .$$

Since $\underline{\underline{h}}$ is symmetric, by using the definitions of (7a), we may write S as

$$(\underline{\underline{\mu}} - \underline{\underline{m}}'')^t (\underline{\underline{h}} n'') (\underline{\underline{\mu}} - \underline{\underline{m}}'') - \underline{\underline{m}}''^t (\underline{\underline{h}} n'') \underline{\underline{m}}'' + \underline{\underline{m}}'^t (\underline{\underline{h}} n') \underline{\underline{m}}' + \underline{\underline{m}}^t (\underline{\underline{h}} n) \underline{\underline{m}} \quad .$$

Now, since

$$\underline{\underline{m}}'^t (\underline{\underline{h}} n') \underline{\underline{m}}' + \underline{\underline{m}}^t (\underline{\underline{h}} n) \underline{\underline{m}} - \underline{\underline{m}}''^t (\underline{\underline{h}} n'') \underline{\underline{m}}'' = \text{tr } \underline{\underline{h}} [n' (\underline{\underline{m}}' \underline{\underline{m}}'^t) + n (\underline{\underline{m}} \underline{\underline{m}}^t) - n'' (\underline{\underline{m}}'' \underline{\underline{m}}''^t)] \quad ,$$

by defining v'' as in (7b), $\underline{v}''$ and $\underline{v}^{*''}$ as in (7c), we may write the kernel (7d) as

$$e^{-\frac{1}{2}(\underline{\mu}-\underline{m}'')^t(\underline{h}n'')(\underline{\mu}-\underline{m}'')} |\underline{h}|^{\frac{1}{2}\delta''} e^{-\frac{1}{2}\text{tr } \underline{h} \underline{v}^{*''}} |\underline{h}|^{\frac{1}{2}v''-1}$$

which is the kernel of Normal-Wishart density with parameter $(\underline{m}'', \underline{v}'', n'', v'')$.

We remark here that a prior of the form (6a) lacks flexibility when v is small because of the manner in which this functional form interrelates the distribution of $\underline{\mu}$ and $\underline{h}$. The nature of this interrelationship is currently being examined, and will be reported in a later paper.[†]

2. Sampling Distributions with Fixed n

We assume here that a sample of size n is to be drawn from an r -dimensional Independent Multinormal process whose parameter $(\underline{\mu}, \underline{h})$ is a random variable having a Normal-Wishart distribution with parameter $(\underline{m}', \underline{v}', n', v')$.

2.1 Conditional Joint Distribution of $(\underline{\tilde{m}}, \underline{\tilde{v}}|\underline{\mu}, \underline{h})$

The conditional joint distribution of the statistic $(\underline{\tilde{m}}, \underline{\tilde{v}})$ given that the process parameter has value $(\underline{\mu}, \underline{h})$ is, provided $v > 0$,

$$D(\underline{\tilde{m}}, \underline{\tilde{v}}|\underline{\mu}, \underline{h}, v) = f_N^{(r)}(\underline{\tilde{m}}|\underline{\mu}, \underline{h}n) f_W^{(r)}(\underline{\tilde{v}}|\underline{h}, v) \quad (8)$$

as shown in section 1.

2.2 Siegel's Generalized Beta Function

Siegel [6] established a class of integral identities with matrix argument that generalize the Beta and Gamma functions. We will use these integral identities in the proofs that the unconditional sampling distributions of $\underline{\tilde{m}}$, of $\underline{\tilde{v}}$ and of $(\underline{\tilde{m}}, \underline{\tilde{v}})$ are as shown in sections 2.3, and 2.4, and 2.5. (In fact, the integrand in Siegel's identity for the generalized Gamma function is the kernel of a Wishart density.)

[†]Ando, A. and Kaufman, G., Extended Natural Conjugate Distributions for the Multinormal Process.

Let $\underline{X}$ be $(r \times r)$ and define

$$\Gamma_r(a) = \pi^{r(r-1)/4} \Gamma(a) \Gamma(a-\frac{1}{2}) \dots \Gamma(a-\frac{r-1}{2}) \quad (9a)$$

$$B_r(a, b) = \frac{\Gamma_r(a) \Gamma_r(b)}{\Gamma_r(a+b)} \quad (9b)$$

where $a > (r-1)/2$, $b > (r-1)/2$. Siegel established the following integral identities: letting $R_X \equiv \{\underline{X} | \underline{X} \text{ is PDS}\}$,

$$\int_{R_X} \frac{|\underline{X}|^{a-\frac{1}{2}(r+1)}}{|\underline{I}+\underline{X}|^{a+b}} d\underline{X} = B_r(a, b) \quad (9c)$$

Defining $\underline{Y} = (\underline{I} + \underline{X})^{-1} \underline{X}$, letting $\underline{V}$ and $\underline{B}$ be real symmetric matrices and letting $\underline{V} < \underline{Y} < \underline{B}$ denote the set $\{\underline{Y} | \underline{B} - \underline{Y}, \underline{Y} - \underline{V} \text{ are PDS}\}$,

$$\int_{R_Y} |\underline{Y}|^{a-\frac{1}{2}(r+1)} |\underline{I}-\underline{Y}|^{b-\frac{1}{2}(r+1)} d\underline{Y} = B_r(a, b) \quad (9d)$$

where the domain of integration $R_Y \equiv \{\underline{Y} | 0 < \underline{Y} < \underline{I}\}$.

We shall define the standardized generalized Beta density function as

$$f_{B*}^{(r)}(\underline{Y} | a, b) \equiv B_r^{-1}(a, b) \frac{|\underline{Y}|^{a-\frac{1}{2}(r+1)} |\underline{I}-\underline{Y}|^{b-\frac{1}{2}(r+1)}}{|\underline{I}+\underline{Y}|^{a+b}}, \quad (9e)$$

$$a > \frac{1}{2}(r-1),$$

$$b > \frac{1}{2}(r-1),$$

$$\underline{Y} \in R_Y.$$

Similarly the standardized inverted generalized Beta density function is defined as

$$f_{iB*}^{(r)}(\underline{X} | a, b) \equiv B_r^{-1}(a, b) \frac{|\underline{X}|^{a-\frac{1}{2}(r+1)}}{|\underline{I}+\underline{X}|^{a+b}}, \quad (9f)$$

$$a > \frac{1}{2}(r-1),$$

$$b > \frac{1}{2}(r-1),$$

$$\underline{X} \in R_X \equiv \{\underline{X} | \underline{X} \text{ is PDS}\}.$$

We also define the non-standardized inverted generalized Beta density function

$$f_{iB}^{(r)}(\underline{V} | a, b, \underline{C}) \equiv [B_r^{-1}(a, b) |\underline{C}|^b] \frac{|\underline{V}|^{a-\frac{1}{2}(r+1)}}{|\underline{V}+\underline{C}|^{a+b}},$$

$$a > \frac{1}{2}(r-1), \quad (9g)$$

$$b > \frac{1}{2}(r-1),$$

$$\underline{C} \text{ is PDS},$$

$$\underline{V} \in R_{\underline{V}} \equiv \{\underline{V} | \underline{V} \text{ is PDS}\}.$$

The functions (9e) and (9f) are related via the integrand transform $\underline{Y} = (\underline{I} + \underline{X})^{-1} \underline{X}$, which has Jacobian $J(\underline{Y}, \underline{X}) = |\underline{I} - \underline{Y}|^{-(r+1)}$. The function (9g) may be transformed into (9f) by an integrand transform $\underline{T} \underline{X} \underline{T}^t = \underline{V}$, where $\underline{T}$ is a non-singular upper triangular matrix such that $\underline{T} \underline{T}^t = \underline{C}$.

2.3 Unconditional Joint Distribution of $(\underline{\tilde{m}}, \underline{\tilde{y}})$

The unconditional (with respect to $(\underline{\tilde{\mu}}, \underline{\tilde{h}})$) joint distribution of $(\underline{\tilde{m}}, \underline{\tilde{y}})$ has density

$$D(\underline{m}, \underline{V} | \underline{m}', \underline{V}', n', v'; n, v)$$

$$= \int_{R_{\underline{\mu}}} \int_{R_{\underline{h}}} f_N^{(r)}(\underline{m} | \underline{\mu}, \underline{h}, n) f_W^{(r)}(\underline{V} | \underline{h}, v) f_{NW}^{(r)}(\underline{\mu}, \underline{h} | \underline{m}', \underline{V}', n', v') d\underline{\mu} d\underline{h} \quad (10a)$$

where the domain of integration $R_{\underline{\mu}}$ of $\underline{\mu}$ is $(-\infty, +\infty)$ and $R_{\underline{h}}$ of $\underline{h}$ is $\{\underline{h} | \underline{h} \text{ is PDS}\}$.

It follows that, provided $v > 0$, $v' > 0$, and $n' > 0$,

$$D(\underline{m}, \underline{V} | \underline{m}', \underline{V}', n', v'; n, v) \propto \frac{|\underline{V}|^{\frac{1}{2}v-1}}{|\underline{V}+\underline{C}|^{\frac{1}{2}(v''+r-1)}} \quad (10b)$$

where

$$n_u = \frac{n' n}{n' + n}, \quad \text{and } \underline{C} = n_u (\underline{m} - \underline{m}') (\underline{m} - \underline{m}')^t + \underline{V}' \quad (10c)$$

Proof: Using (4a') and (8) we may write

$$D(\underline{m}, \underline{v} | \underline{m}', n', v'; n, v) \\ = \int_{\underline{R}_h} \left[\int_{\underline{R}_u} f_N^{(r)}(\underline{m} | \underline{u}, \underline{h}n) f_N^{(r)}(\underline{u} | \underline{m}', \underline{h}n') d\underline{u} \right] \cdot f_W^{(r)}(\underline{v} | \underline{h}, v) f_W^{(r)}(\underline{h} | \underline{v}', v') d\underline{h}.$$

The inner integral is $f_N^{(r)}(\underline{m} | \underline{m}', \underline{h}n_u)$. Hence the total integral may be written as

$$\int_{\underline{R}_h} f_N^{(r)}(\underline{m} | \underline{m}', \underline{h}n_u) f_W^{(r)}(\underline{v} | \underline{h}, v) f_W^{(r)}(\underline{h} | \underline{v}', v') d\underline{h}$$

which upon replacing the f_N and f_W 's by their respective formulas and dropping constants becomes

$$|\underline{v}|^{\frac{1}{2}v-1} \int_{\underline{R}_h} e^{-\frac{1}{2}n_u(\underline{m}-\underline{m}')^t \underline{h}(\underline{m}-\underline{m}')} \cdot e^{-\frac{1}{2}tr \underline{h}(\underline{v}'+\underline{v})} |\underline{h}|^{\frac{1}{2}(v'+v+r+\delta+\delta'-\delta''-\phi-1)-1} d\underline{h}$$

Using the definitions of (7), this equals

$$|\underline{v}|^{\frac{1}{2}v-1} \int_{\underline{R}_h} e^{-\frac{1}{2}tr \underline{h}\{n_u(\underline{m}-\underline{m}')(\underline{m}-\underline{m}')^t + \underline{v}'+\underline{v}\}} |\underline{h}|^{\frac{1}{2}v''-1} d\underline{h}.$$

Letting $\underline{B} = n_u(\underline{m}-\underline{m}')(\underline{m}-\underline{m}')^t + \underline{v}'+\underline{v}$ from (4c) we see that the integrand in the above integral is, aside from the multiplicative constant $w(r, v'') |\underline{B}|^{\frac{1}{2}(v''+r-1)}$, a

Wishart density with parameter $(\underline{B}, v'')$. Hence, apart from a normalizing constant depending on neither $\underline{v}$ nor $\underline{B}$,

$$D(\underline{m}, \underline{v} | \underline{m}', \underline{v}', n', v'; n, v) \propto \frac{|\underline{v}|^{\frac{1}{2}v-1}}{|\underline{B}|^{\frac{1}{2}(v''+r-1)}} = \frac{|\underline{v}|^{\frac{1}{2}v-1}}{|\underline{v}+\underline{v}'+n_u(\underline{m}-\underline{m}')(\underline{m}-\underline{m}')^t|^{\frac{1}{2}(v''+r-1)}}$$

proving (10b).

2.4 Unconditional Distribution of $\tilde{\underline{m}}$

The kernel of the unconditional (with respect to $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$ distribution of $\tilde{\underline{m}}$ can be found three ways: by utilizing the fact that $\tilde{\underline{m}}|\tilde{\underline{\mu}}$ and $\tilde{\underline{\mu}}$ are conditionally independent given $\tilde{\underline{h}}=\underline{h}$ and then finding the unconditional distribution of $\tilde{\underline{m}}$ as regards $\tilde{\underline{h}}$. By integrating $D(\underline{m}, \underline{v}|\underline{m}', n', v'; n, v)$ over the range of $\underline{v}$ when this distribution exists i.e. over $R_{\underline{v}} \equiv \{\underline{v}|\underline{v} \text{ is PDS}\}$; or we may find it by integrating the kernel of the marginal likelihood of $\tilde{\underline{m}}$ defined in (4b) multiplied by the kernel of the prior density of $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$ over $R_{\underline{\mu}}$ and $R_{\underline{h}}$. The first and third methods have the merit of in no way depending on whether or not $\underline{v}$ is singular-- that is, even if $v \leq 0$ ($n \leq r$), the proofs go through, which of course is not the case if we proceed according to the second method. We show first by the second method that when $v > 0$, $v' > 0$, and $\underline{v}^* = \underline{v}'$ is PDS,

$$D(\underline{m}|\underline{m}', \underline{v}', n', v'; n, v) = \int_{R_{\underline{v}}} D(\underline{m}, \underline{v}|\underline{m}', \underline{v}', n', v'; n, v) d\underline{v} \quad (11a)$$

$$\propto [1 + n_u(\underline{m}-\underline{m}')^t \underline{H}_{\underline{v}}^{-1} (\underline{m}-\underline{m}')]^{-\frac{1}{2}(v'+r)}$$

We then show by the first method that this result holds even when $v \leq 0$. If we define $\underline{H}_{\underline{v}} = v' n_u \underline{v}^{-1}$, then (11a) may be rewritten as

$$[v' + (\underline{m}-\underline{m}')^t \underline{H}_{\underline{v}} (\underline{m}-\underline{m}')]^{-\frac{1}{2}(v'+r)} \quad (11b)$$

Provided $v' > 0$, $n_u > 0$ and $\underline{H}_{\underline{v}}$ is PDS this is the kernel of the nondegenerate Student density function with parameter $(\underline{m}', \underline{H}_{\underline{v}}, v')$ as defined in formula (8-26) of [5].

Proof: In (10b) let $a = (\frac{1}{2}v-1) + \frac{1}{2}(r+1)$ and $b = \frac{1}{2}(v'+r-1) - a = \frac{1}{2}(v'+r)$. Then the kernel of the unconditional distribution of $\tilde{\underline{m}}$ is proportional to

$$R_{\underline{v}} = \int \cdot \frac{|\underline{v}|^{a-\frac{1}{2}(r+1)}}{|\underline{v}+\underline{c}|^{a+b}} d\underline{v} \quad , \quad (12)$$

where $R_{\underline{v}} = \{\underline{v} | \underline{v} \text{ is PDS}\}$. Then by (9g),

$$D(\underline{m} | \underline{m}', \underline{v}', n'; n) \propto |n_u(\underline{m}-\underline{m}')(\underline{m}-\underline{m}')^t + \underline{v}'|^{-b} = (1 + (\underline{m}-\underline{m}')^t n_u \underline{v}'^{-1} (\underline{m}-\underline{m}'))^{-b} \quad ,$$

establishing (11a).

Alternate Proof: Another way of establishing (11) is to use the fact that the kernel of the marginal likelihood of $\tilde{\underline{m}}$ given for $n > 0$ observations the parameter $(\underline{\mu}, \underline{h})$ is by (5) and (4b') whether or not $v \leq 0$

$$e^{-\frac{1}{2}n(\underline{m}-\underline{\mu})^t \underline{h}(\underline{m}-\underline{\mu})} |\underline{h}|^{\frac{1}{2}} \quad . \quad (4b)$$

Furthermore, conditional on $\tilde{\underline{h}} = \underline{h}$, $\tilde{\underline{\mu}}$ and $\tilde{\underline{\epsilon}} \equiv \tilde{\underline{m}} - \tilde{\underline{\mu}}$ are independent Normal random vectors; and so $\tilde{\underline{m}} = \tilde{\underline{\mu}} + \tilde{\underline{\epsilon}}$ is Normal with mean vector

$$E(\tilde{\underline{m}}) = E(\tilde{\underline{\mu}}) + E(\tilde{\underline{\epsilon}}) = \underline{m}' + \underline{0} = \underline{m}'$$

and variance-covariance matrix

$$V(\tilde{\underline{m}}) = V(\tilde{\underline{\mu}}) + V(\tilde{\underline{\epsilon}}) = (\underline{h} \ n_u)^{-1} \quad .$$

Thus integrating with respect to $\underline{h}$,

$$D(\underline{m} | \underline{v}', n', v'; n, v) = \int_{R_h} e^{-\frac{1}{2} \text{tr } \underline{h} \{n_u(\underline{m}-\underline{m}')(\underline{m}-\underline{m}')^t + \underline{v}'\}} |\underline{h}|^{\frac{1}{2}(v'+1)-1} d\underline{h}$$

As the integrand in the above integral is the kernel of a Wishart density with parameter $(n_u(\underline{m}-\underline{m}')(\underline{m}-\underline{m}')^t + \underline{v}', v'+1)$,

$$D(\underline{m} | \underline{v}', n', v'; n, v) \propto |n_u(\underline{m}-\underline{m}')(\underline{m}-\underline{m}')^t + \underline{v}'|^{-\frac{1}{2}(v'+r)}$$

and (11) follows directly.

2.5 Unconditional Distribution of $\tilde{\underline{v}}$ When $v > 0$

The kernel of the unconditional distribution of $\tilde{\underline{v}}$ can be found by integrating $D(\underline{m}, \underline{v} | \underline{m}', \underline{v}', n', v'; n, v)$ over the range of $\underline{m}$, or by integrating the product of the kernel of the marginal likelihood (4c) of $\tilde{\underline{v}}$ and the kernel of the distribution (5a) of $(\underline{\mu}, \underline{h})$ with parameter $(\underline{m}', \underline{v}', n', v')$ over the range of $(\underline{\mu}, \underline{h})$. We show by the former method that for $a > \frac{1}{2}(r-1)$ and $b > \frac{1}{2}(r-1)$, when $v > 0$ and $\underline{v}$ is PDS

$$D(\underline{v} | \underline{m}', \underline{v}', n'; n) \propto \frac{|\underline{v}|^{a-\frac{1}{2}(r+1)}}{|\underline{v}'+\underline{v}|^{a+b}} \quad (16a)$$

where $a = \frac{1}{2}(v+r-1)$ and $b = \frac{1}{2}(v'+r-1)$. Letting $\underline{K}$ be a non-singular upper triangular matrix such that $\underline{K} \underline{K}^t = \underline{v}'$ we may make the integrand transform $\underline{K} \underline{Z} \underline{K}^t = \underline{v}$ and write (16a) as

$$D(\underline{Z} | \underline{m}', \underline{v}', n'; n) \propto \frac{|\underline{Z}|^{a-\frac{1}{2}(r+1)}}{|\underline{I}+\underline{Z}|^{a+b}} \quad (16b)$$

Formula (16b) is the kernel of a standardized inverted generalized Beta function with parameter (a, b) .

Proof: From (10b)

$$D(\underline{m}, \underline{v} | \underline{m}', \underline{v}', n'; n, v) \propto |\underline{v}|^{\frac{1}{2}v-1} |\underline{v}'+\underline{v}+n_u(\underline{m}-\underline{m}')(\underline{m}-\underline{m}')^t|^{-\frac{1}{2}(v''+r-1)}$$

Conditional on $\tilde{\underline{V}} = \underline{V}$, the second determinant is the kernel of a Student density of $\underline{m}$ with parameter $(\underline{m}, n_u(v''-1)[\underline{V}' + \underline{V}], v''-1)$. That part of the constant which normalizes this Student density and which involves $\underline{V}$ is $|\underline{V}' + \underline{V}|^{-\frac{1}{2}(v''+r-2)}$; hence (16a) follows.

Now since $\underline{V}'$ is PDS there is a non-singular triangular matrix $\underline{K}$ of order r such that $\underline{K} \underline{K}^t = \underline{V}'$. If we make the integrand transform $\underline{K} \underline{Z} \underline{K}^t = \underline{V}'$, and let $J(\underline{Z}, \underline{V})$ denote the Jacobian of the transform, the transformed kernel may then be written as

$$\frac{|\underline{K} \underline{Z} \underline{K}^t|^{a-\frac{1}{2}(r+1)}}{|\underline{K} \underline{K}^t + \underline{K} \underline{Z} \underline{K}^t|^{a+b}} \cdot J(\underline{Z}, \underline{V}) = |\underline{K} \underline{K}^t|^{-b-\frac{1}{2}(r+1)} J(\underline{Z}, \underline{V}) \cdot \frac{|\underline{Z}|^{a-\frac{1}{2}(r+1)}}{|\underline{I} + \underline{Z}|^{a+b}}.$$

Since $J(\underline{Z}, \underline{V}) = |\underline{K}|^{r+1} = |\underline{K} \underline{K}^t|^{\frac{1}{2}(r+1)} = |\underline{V}'|^{\frac{1}{2}(r+1)}$, we may write the transformed kernel as shown in (16b).

3. Preposterior Analysis with Fixed $n > 0$

We assume that a sample of fixed size $n > 0$ is to be drawn from an r -dimensional Multinormal process whose mean vector $\underline{\mu}$ and matrix precision $\underline{h}$ are not known with certainty but are regarded as random variables $(\underline{\mu}, \underline{h})$ having a prior Normal-Wishart distribution with parameter $(\underline{m}', \underline{V}', n', v')$ where $n' > 0$, $v' > 0$, but $\underline{V}'$ may or may not be singular.

3.1 Joint Distribution of $(\underline{m}'', \underline{V}'')$

The joint density of $(\underline{m}'', \underline{V}'')$ is, provided $n' > 0$, $v' > 0$, and $v > 0$ is

$$D(\underline{m}'', \underline{V}'' | \underline{m}', \underline{V}', n', v'; n, v) \propto \frac{|\underline{V}'' - \underline{V}' - n^*(\underline{m}'' - \underline{m}')(\underline{m}'' - \underline{m}')^t|^{\frac{1}{2}v-1}}{|\underline{V}''|^{\frac{1}{2}(v''+r)}} \quad (17a)$$

where

$$n^* = n' n'' / n.$$

(17b)

and the range of $(\underline{m}'', \underline{v}'')$ is

$$R(\underline{m}'', \underline{v}'') \equiv \{(\underline{m}'', \underline{v}'') \mid -\infty < \underline{m}'' < +\infty \text{ and } \underline{v}'' - \underline{c} \text{ is PDS}\} \quad (17c)$$

Proof: Following the argument of subsections 12.1.4 and 12.6.1 of [5] we can establish that

$$n_u(\underline{m} - \underline{m}')^t h(\underline{m} - \underline{m}') = n^*(\underline{m}'' - \underline{m}')^t h(\underline{m}'' - \underline{m}') \quad (18a)$$

where $n^* = n' n'' / n$. The same line of argument allows us to write

$$\underline{v}'' = \underline{v}' + \underline{v} + n'(\underline{m}' \underline{m}')^t + n(\underline{m} \underline{m})^t - n''(\underline{m}'' \underline{m}')^t = \underline{v}' + \underline{v} + n^*(\underline{m}'' - \underline{m}')(\underline{m}'' - \underline{m}')^t. \quad (18b)$$

From (7) we have

$$(\underline{m}, \underline{v}) = \left(\frac{1}{n} (n'' \underline{m}'' - n' \underline{m}'), \underline{v}'' - \underline{v}' - n^*(\underline{m}'' - \underline{m}')(\underline{m}'' - \underline{m}')^t \right).$$

Letting $J(\underline{m}'', \underline{v}''; \underline{m}, \underline{v})$ denote the Jacobian of the integrand transformation from $(\underline{m}, \underline{v})$ to $(\underline{m}'', \underline{v}'')$, we make this transformation in (10b), obtaining (17a). Since $J(\underline{m}'', \underline{v}''; \underline{m}, \underline{v}) = J(\underline{m}'', \underline{m}) J(\underline{v}'', \underline{v})$ and both $J(\underline{m}'', \underline{m})$ and $J(\underline{v}'', \underline{v})$ are constants involving neither $\underline{m}'$ nor $\underline{v}'$, neither does $J(\underline{m}'', \underline{v}''; \underline{m}, \underline{v})$. When $v \leq 0$ and $\underline{v}$ is singular the numerator in (17a) vanishes, so that the density exists only if $v > 0$. However if $v > 0$, the kernel in (17a) exists even if $\underline{v}'$ is singular ($\underline{v}' \neq 0$). Hence we write the kernel as shown in (17a).

That the range of $(\underline{m}'', \underline{v}'')$ is $R(\underline{m}'', \underline{v}'')$ as defined in (17c) follows directly from the definitions of $\underline{m}'$ and $\underline{v}''$, of $R_{\underline{m}}$ and $R_{\underline{v}}$, and of $\underline{m}'$ and $\underline{v}'$.

3.2 Some Distributions of $\underline{\tilde{m}}''$

The unconditional distribution of $\underline{\tilde{m}}''$ is easily derived from the unconditional distribution (11b) of $\underline{\tilde{m}}$: provided $n > 0$, $n' > 0$, $v' > 0$, and $\underline{v}'$ is PDS,

$$D(\underline{m}'' \mid \underline{m}', \underline{v}', n', v'; n, v) = f_S^{(r)}(\underline{m}'' \mid \underline{m}', (n''/n)^2 H_{\underline{v}}, v') \quad (19a)$$

where

$$H_{\underline{v}} = v' n_u \underline{v}'^{-1}, \quad (19b)$$

and the conditional distribution of $\underline{\tilde{m}}'$ given $\underline{\tilde{V}}'' = \underline{V}''$ is, provided $v > 0$, $n' > 0$, $v' > 0$,

$$D(\underline{m}'' | \underline{m}', \underline{V}', n', v'; n, v, \underline{V}'') = f_{IS}^{(r)}(\underline{m}'' | \underline{m}', \underline{H}_*, v) \quad (20a)$$

where

$$\underline{H}_* = vn^*(\underline{V}'' - \underline{V}')^{-1} \quad (20b)$$

The right hand side of (20a) is the inverted Student density function with parameter $(\underline{m}', \underline{H}_*, v)$ as defined in formula (8-34) of [5].

Proof: Since $\underline{m}'' = \frac{1}{n''} (n' \underline{m}' + n \underline{m})$ and since from (11b) when $n > 0$, $n' > 0$, $v' > 0$, $\underline{V}'$ is PDS, and

$$\underline{\tilde{m}} \sim f_S^{(r)}(\underline{m} | \underline{m}', \underline{H}_v, v'),$$

by Theorem 1 of subsection 8.3.2 of [5],

$$\underline{\tilde{m}}'' \sim f_S^{(r)}(\underline{m}'' | \underline{m}', (n''/n)^2 \underline{H}_v, v').$$

To prove (20) observe that the kernel of the conditional distribution of $\underline{\tilde{m}}''$ given $\underline{\tilde{V}}'' = \underline{V}''$ is proportional to (17a), and so

$$D(\underline{m}'' | \underline{m}', \underline{V}', n', v'; n, v, \underline{V}'') \propto |\underline{V}'' - \underline{V}' - n^*(\underline{m}'' - \underline{m}')(\underline{m}'' - \underline{m}')^t|^{\frac{1}{2}v-1}.$$

Since

$$\underline{V}'' - \underline{V}' = \underline{V} + n^*(\underline{m}'' - \underline{m}')(\underline{m}'' - \underline{m}')^t,$$

$(\underline{V}'' - \underline{V}')$ will be PDS, as long as $v > 0$, and so $(\underline{V}'' - \underline{V}')^{-1}$ is also PDS. Using a well known

determinantal identity and letting $\underline{H}_*$ be as defined in (20b), when $\underline{V}'' - \underline{V}'$

is PDS we may write the density of $\underline{\tilde{m}}''$ as

$$[1 - n^*(\underline{m}'' - \underline{m}')^t (\underline{V}'' - \underline{V}')^{-1} (\underline{m}'' - \underline{m}')]^{\frac{1}{2}v-1} = v^{-\frac{1}{2}v+1} [v - (\underline{m}'' - \underline{m}')^t \underline{H}_* (\underline{m}'' - \underline{m}')]^{\frac{1}{2}v-1},$$

which, aside from the constant $v^{-\frac{1}{2}v+1}$, is the kernel of an inverted Student density function with parameter $(\underline{m}', \underline{H}_*, v)$.

3.3 Analysis When $n \leq r$

Even when $n \leq r$, it is possible to do Bayesian inference on $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$ by appropriately structuring the prior so that the posterior of $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$ is non-degenerate.

For example, if the data generating process is Multinormal, if we assign a prior on $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$ with parameter $(\underline{0}, \underline{0}, \underline{0}, \underline{0})$, and then observe a sample $\underline{x}^{(1)}, \dots, \underline{x}^{(n)}$ where $n \leq r$, then $v \leq 0$ and the posterior parameters defined in (7) assume values

$$n''=n, \quad v''=v+r-\phi-1=0, \quad \underline{m}''=\underline{m}, \quad \underline{v}''=\underline{0}.$$

The posterior distribution of $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$ is degenerate under these circumstances.

If, however, we insist on assigning a very diffuse prior on $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$, but are willing to introduce just enough prior information to make $\underline{v}''$ non-singular and $v'' > 0$ then the posterior distribution is non-degenerate; e.g. assign $v'=1$, $\underline{v}'=M \underline{I}$, $M \gg 0$, and leave $n'=0$ and $\underline{m}'=\underline{0}$, so that $v''=1$, $\underline{v}''=\underline{v}+\underline{v}'=\underline{v}+M\underline{I}$. In this case we have a bona fide non-degenerate Normal-Wishart posterior distribution of $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$.

Furthermore, the unconditional distribution of the next sample observation $\underline{x}^{(n+1)}$ exists and is, by (11b), multivariate Student with parameter

$$(\underline{m}, (\frac{n}{n+1})(\underline{v}+M\underline{I})^{-1}, 1).$$

In addition, for this example the distribution, unconditional as regards $(\tilde{\underline{\mu}}, \tilde{\underline{h}})$, of the mean $\tilde{\underline{x}}$ of the next n^0 observations exists even though $n \leq r$ and is by (11b) multivariate Student with parameter $(\underline{m}, n_u(\underline{v}+M\underline{I})^{-1}, 1)$, where $n_u = n^0 n / (n^0 + n)$. This distribution is, in effect, a probabilistic forecast of $\tilde{\underline{x}}$. From (19a) it also follows that the distribution, unconditional as regards

$(\tilde{\underline{\mu}}, \tilde{\underline{h}})$, of the posterior mean $\tilde{\underline{m}}$ prior to observing $\underline{x}^{(n+1)}, \dots, \underline{x}^{(n+n^0)}$ is multivariate Student with parameter $(\underline{m}, (1 + \frac{n}{n_0})n\underline{MI}, 1)$.

REFERENCES

- [1] Anderson, T.W., Introduction to Multivariate Statistical Analysis, (John Wiley and Sons, Inc., New York, 1958.)
- [2] Deemer, W.L. and Olkin, I., "The Jacobians of Certain Matrix Transformations Useful in Multivariate Analysis, Biometrika, Vol. 40, pp. 43-46 (1953).
- [3] Geisser, S. and Cornfield, J., "Posterior Distributions for Multivariate Normal Parameters," Journal of the Royal Statistical Society, Vol. 25, No. 2, 1963.
- [4] Jeffreys, H., Theory of Probability (Oxford University Press, Amen House, London 1961).
- [5] Raiffa, H. and Schlaifer, R., Applied Statistical Decision Theory (Division of Research, Harvard Business School, Boston 1961).
- [6] Siegel, C.L., "Uber die analytische theorie der quadratischen Formen," Annals of Mathematics, Vol. 36, pp. 527-606 (1935).
- [7] Zellner, A. and Tiao, G.C., "On the Bayesian Estimation of Multivariate Regression," Journal of the Royal Statistical Society, (Series B), Vol. 26, Part 2, 1964.

